



Distinct evolutionary trajectories of MHC class I and class II genes in Old World finches and buntings

Piotr Minias ¹ · Radosław Włodarczyk ¹ · Magdalena Remisiewicz ² · Ioana Cobzaru ³ · Tomasz Janiszewski¹

Received: 28 October 2020 / Revised: 8 March 2021 / Accepted: 15 March 2021 / Published online: 6 April 2021
© The Author(s), under exclusive licence to The Genetics Society 2021

Abstract

Major histocompatibility complex (MHC) genes code for key proteins of the adaptive immune system, which present antigens from intra-cellular (MHC class I) and extra-cellular (MHC class II) pathogens. Because of their unprecedented diversity, MHC genes have long been an object of scientific interest, but due to methodological difficulties in genotyping of duplicated loci, our knowledge on the evolution of the MHC across different vertebrate lineages is still limited. Here, we compared the evolution of MHC class I and class II genes in three sister clades of common passerine birds, finches (Fringillinae and Carduelinae) and buntings (Emberizidae) using a uniform methodological (genotyping and data processing) approach and uniform sample sizes. Our analyses revealed contrasting evolutionary trajectories of the two MHC classes. We found a stronger signature of pervasive positive selection and higher allele diversity (allele numbers) at the MHC class I than class II. In contrast, MHC class II genes showed greater allele divergence (in terms of nucleotide diversity) and a much stronger recombination (gene conversion) signal. Gene copy numbers at both MHC class I and class II evolved via fluctuating selection and drift (Brownian Motion evolution), but the evolutionary rate was higher at class I. Our study constitutes one of few existing examples, where evolution of MHC class I and class II genes was directly compared using a multi-species approach. We recommend that re-focusing MHC research from single-species and single-class approaches towards multi-species analyses of both MHC classes can substantially increase our understanding MHC evolution in a broad phylogenetic context.

These authors contributed equally: Piotr Minias, Radosław Włodarczyk

Associate editor: Barbara Mable

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00427-8>.

✉ Piotr Minias
pminias@op.pl

- ¹ Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Łódź, Poland
- ² Bird Migration Research Station, Faculty of Biology, University of Gdańsk, Gdańsk, Poland
- ³ Institute of Biology Bucharest, Romanian Academy, Bucharest, Romania

Introduction

Genes of the major histocompatibility complex (MHC) constitute a key component of the adaptive immune system in vertebrates, as they code for proteins directly involved in the presentation of intra-cellular and extra-cellular pathogens (MHC class I and class II, respectively). Although MHC molecules of both classes differ in their basic structure, they share a notable feature—a special groove that binds antigens (the so-called peptide-binding region, PBR). MHC-bound antigens (or self-peptides) are presented to T lymphocytes and may be recognized by T-cell receptors (TCRs), which triggers a cascade of immune responses that are aimed at pathogen elimination (Janeway et al. 2001).

In general, the evolution of the MHC in jawed vertebrates proceeded from a simple design with relatively few genes in Chondrichthyes fish to a much more complex architecture in mammals (Kulski et al. 2002), where duplicated MHC class I and class II gene copies are scattered over several loci (defined as specific positions on the chromosome where gene copies are located). In birds,

ancestral phylogenetic lineages have relatively low numbers of MHC gene copies, e.g., most species from one of the oldest avian clades (order: Galliformes) have only one dominantly expressed classical MHC gene copy at each class (Kaufman et al. 1999). The MHC duplication rate was generally suppressed in non-passerine birds, but greatly accelerated during the passerine (order: Passeriformes) radiation (Minias et al. 2019). Among oscine passerines, the highest numbers of MHC copies were recorded in the Sylvioidea (MHC class I) and Passeroidea (MHC class II) superfamilies (Minias et al. 2019), reaching over 30 copies in some species (Sepil et al. 2012; Biedrzycka et al. 2017a). At the inter-specific level, different species may have different evolutionary optima of MHC copy numbers and some copies may be deleted or inactivated via deleterious mutations (turned into pseudogenes) over evolutionary times, consistent with the birth-and-death model of MHC evolution (Nei et al. 1997). The presence of multiple MHC copies (generated through gene duplications) is thought to confer fitness benefits, as it is expected to increase the number of alleles (unique gene variants) expressed within individuals and, thus, should enhance the spectrum of pathogens recognized by an individual organism (Bentkowski and Radwan 2019). On the other hand, the number of MHC copies cannot expand in an unlimited manner and, according to the optimality hypothesis (Nowak et al. 1992), copy numbers are thought to be restricted by the inherent costs of allele expression, e.g., depletion of TCR repertoire, as recently shown in rodents (Migalska et al. 2019).

One of the most notable features of the MHC is an extreme level of allelic diversity (the number of alleles per gene) within populations, which is maintained by various forms of pathogen-driven balancing selection (Spurgin and Richardson 2010; Radwan et al. 2020), including the mechanisms of overdominant selection (heterozygote advantage, i.e., higher fitness of heterozygous over homozygous genotypes), negative frequency-dependent selection (higher fitness of rare genotypes) and fluctuating selection (spatial and temporal variation in fitness value of particular genotypes) (reviewed in Radwan et al. 2020). Although the relative roles of these three mechanisms in generating MHC diversity are very difficult to separate (Radwan et al. 2020), their combined effect can generate and maintain thousands of MHC alleles within natural vertebrate populations (e.g., Biedrzycka et al. 2017a). Also, strong balancing selection can maintain highly adaptive alleles for long evolutionary periods and, consequently, coalescent time of MHC alleles may span many millions of years (Takahata 1990). Persistence of MHC alleles beyond species diversification produces a specific phylogenetic pattern where some alleles are more similar between species than within species (so-called trans-species polymorphism) (Takahata 1990).

So far, the evolution of the mammalian MHC has been most rigorously studied among all vertebrate groups and the major evolutionary features of these genes were first revealed during the pioneering research on mammals (Hess and Edwards 2002; Bernatchez and Landry 2003). However, the patterns of molecular evolution of the MHC are thought to largely differ between mammals and birds. The mammalian MHC primarily evolves according to the divergent evolution model (Ota and Nei 1994), where different loci are maintained independently of each other after duplication events. Consistently with this hypothesis, some MHC loci were found to have duplicated very early in the mammalian radiation and evolved independently since the divergence of marsupial and placental mammals (Takahashi et al. 2000). In contrast, the avian MHC is characterized by more recent duplications with little evidence for orthologous relationships between loci across different lineages (Edwards et al. 1995, but see Burri et al. 2008). Also, different MHC loci in birds are thought to be continuously homogenized (e.g., by gene conversion) and often evolve as a single unit, consistently with the concerted (rather than divergent) model of evolution (Nei and Rooney 2005). Molecular analyses provided evidence for the concerted evolution of MHC class II in several avian lineages (Witzell et al. 1999; Gillingham et al. 2016), but it remains unresolved whether similar mechanism shapes MHC class I diversity in birds.

Although the recent advancement of second- and third-generation sequencing technologies sparked research on the MHC in a broad spectrum of non-model organisms (O'Connor et al. 2019), most studies focus either on a single taxon or on a single MHC class, which seriously limits our knowledge on whether MHC class I and class II genes have similar evolutionary trajectories in the same vertebrate lineages. In birds, comparisons of MHC classes I and II evolution in a multi-species framework have been conducted for several non-passerine clades (Gillingham et al. 2016; Minias et al. 2016; Sallaberry-Pincheira et al. 2016), whereas multi-species analyses of passerine MHC focused either on class I (e.g., O'Connor et al. 2016; Drews et al. 2017) or class II (e.g., Jarvi et al. 2004; Balasubramaniam et al. 2016). On the other hand, the few phylogenetically robust analyses of MHC evolution in birds (across a wide spectrum of avian lineages) have usually been based on publicly available genetic resources (GenBank database), thus combining data generated with a variety of methodological (genotyping and data processing) approaches and suffering from unbalanced sample sizes (both in terms of the number of individuals/species genotyped and species composition) (e.g., Minias et al. 2018, 2019). Although this kind of research can provide some broad pictures of MHC evolution in non-model taxa, due to its inherent methodological limitations it may not be particularly appropriate to

detect any fine-scale differences in the evolutionary mechanisms between MHC class I and class II. In fact, evolutionary trajectories of the two MHC classes should ideally be compared using the same set of taxa, uniform sample sizes and uniform genotyping/data processing methodology, but such analyses are still lacking for most avian lineages, including passerines.

The aim of this study was to compare the evolution of MHC class I and class II genes in oscine passerines using three sister clades of finches (Fringillinae and Carduelinae) and buntings (Emberizidae) as our study object. Despite the fact that both finches and buntings are relatively diverse (201 and 42 species, respectively; Winkler et al. 2015) and belong to the most commonly represented passerine families in the Western Palearctic avifauna, they have been largely under-represented in passerine MHC research and our knowledge on the MHC evolution in these clades is fragmentary and limited to just a couple of species (Arnaiz-Villena et al. 2007; Li et al. 2017; Drews and Westerdahl 2019). For example, phylogenetic analyses of MHC class I in three carduelid finches (*Serinus*) provided evidence for trans-species polymorphism and slower evolution of MHC class I in continental (African and Asian) species compared to the species from Canary Islands (Arnaiz-Villena et al. 2007). Examination of MHC class I expression profiles in another carduelid finch, the Eurasian siskin *Spinus spinus*, allowed identification of classical and nonclassical alleles and revealed that several MHC I genes are highly expressed in this species (Drews and Westerdahl 2019). As far as we are aware, MHC class II has not been studied in Fringillidae, but its polymorphism was assessed in two east Asian *Emberiza* species (Emberizidae), providing support for positive selection signature and trans-species polymorphism, but no evidence for recombination occurring at these genes (Li et al. 2017). To the best of our knowledge, the evolution of MHC class I and class II has never been directly compared within these avian lineages. Here, we reconstructed and compared the evolutionary history of MHC class I and class II genes across the clades of Emberizidae, Carduelinae and Fringillinae using uniform genotyping and data processing methodology, as well as uniform sample sizes. Specifically, we hypothesized that contrasting selective regimes from intra-cellular and extra-cellular pathogens could have generated and maintained differences in: (1) gene copy number variation (duplication rates), (2) sequence polymorphism, (3) recombination (gene conversion), (4) nucleotide substitution rates and (5) phylogenetic patterns (e.g., the level of trans-species polymorphism) between the MHC class I and class II genes. Such comparative analyses of evolutionary trajectories in the two MHC classes provide a unique opportunity to infer how selection from two distinct groups of pathogens (intra-cellular and extra-cellular) affected the evolution on their hosts.

Materials and methods

Sample collection and preparation

The MHC polymorphism was examined in eleven species from three clades: Emberizidae (three species): corn bunting *Emberiza calandra*, yellowhammer *Emberiza citrinella* and reed bunting *Emberiza schoeniclus*; Carduelinae (six species): common redpoll *Acanthis flammea*, European goldfinch *Carduelis carduelis*, European greenfinch *Chloris chloris*, hawfinch *Coccothraustes coccothraustes*, Eurasian bullfinch *Pyrrhula pyrrhula* and Eurasian siskin; and Fringillinae (two species): common chaffinch *Fringilla coelebs* and brambling *Fringilla montifringilla*. The differences in the number of species sampled per clade within our study roughly reflected the true differences in the species richness between the clades, with Fringillinae (3 extant species) and Carduelinae (166 extant species) being the least and the most species rich (Winkler et al. 2015). We captured and sampled blood from eight individuals per species and sampling took place at four locations in Poland (5–53 individuals per site; Table S1 in Appendix 1). All samples were collected in March–December 2018 during breeding, autumn migration and early wintering seasons. Since most individuals (ca. 85%) were captured during migration or wintering, we had no reliable information on their origin and different sampling sites could not be treated as separate populations (birds from the same population could migrate through different sampling sites and birds from different populations could migrate through the same sampling site). Consequently, sampling location was not taken into account in our analyses. Blood was collected from each captured individual by puncturing the ulnar vein with a disposable needle. Ca. 50 µl of blood was collected with a heparinized capillary, transferred into Eppendorf tubes with 95% ethanol and kept at 2 °C until laboratory analyses. Genomic DNA was extracted from blood samples with Bio-Trace DNA Purification Kit (EURx, Gdansk, Poland).

MHC amplification and Illumina sequencing

In non-model species with relatively high allelic diversity of the MHC (as in the case of our study system, see results for details), any single pair of primers is unlikely to amplify all alleles existing within populations (and within individuals). Since different primers can amplify different MHC alleles, a combination of genotyping data obtained with two (or more) independent pairs of primers is likely to provide a more reliable (less underestimated) estimate of the total number of alleles present within individuals and, consequently, of the gene copy number. Thus, each MHC class (I and II) was genotyped using two sets of independent primers (Table 1 and Fig. S1 in Appendix 1),

Table 1 Primers used for MHC class I exon 3 and MHC class II exon 2 genotyping in finches and buntings.

MHC	Exon	Primer pair	Forward primer	Reverse primer	References
Class I	3	PP1_1	MhcPasCI-FW 5'-CSCSCAGGTCTSCACAC-3'	MhcPasCI-RV 5'-CWCARKAATTCTGYTCHCACC-3'	Alcaide et al. (2013)
		PP1_2	HNalla 5'-TCCCCACAGGTCTCCACAC-3'	RV 5'-TGCGCTCCAGCTCCYTCTGCC-3'	O'Connor et al. (2016)
Class II	2	PP2_1	Int1f.7 5'-ATGGGACCCCAAAAGTGATT-3'	Int2r.1 5'-CCGAGGGGACACGCTCT-3'	Aguilar et al. (2006)
		PP2_2	MHCIIPas-E2iF 5'-GAGTGTACTTCATTAACGGCAC-3'	MHCIIPas-E2iR 5'-CYNGTAGTTGTGNCGGCAG-3'	Canal et al. (2010)

which were previously used for MHC genotyping in a wide range of passerine species (Aguilar et al. 2006; Alcaide et al. 2013; Canal et al. 2010; O'Connor et al. 2016). The primers successfully amplified fragments of class I α exon 3 or class II β exon 2 in all our study species and following the approach by O'Connor et al. (2016) we did not aim to design species-specific primers, so that the results were more comparable at the inter-specific level. Although peptide-binding grooves of both MHC classes I and II molecules are coded by two exons (exon 2 and 3 coding for $\alpha 1$ and $\alpha 2$ domains at class I; exon 2 α and 2 β coding for $\alpha 1$ and $\beta 1$ domains at class II; Chen et al. 2015), class I α exon 3 and class II β exon 2 are thought to be most polymorphic and are most commonly targeted while genotyping MHC in non-model vertebrates, including birds (avian MHC sequences deposited in GenBank: 255 exon 2 class I α vs. 6691 exon 3 class I α ; 30 exon 2 class II α vs. 4629 exon 2 class II β ; as accessed on 5 October 2020). For amplifications we used fusion primers composed of the MHC primer (as described in Table 1), a 7-bp barcode indicating sample identity, and Illumina Nextera Transposase adaptor sequences (Illumina Corp., San Diego, CA, USA). All PCR reactions were conducted in a final volume of 20 μ l containing 10 μ l of 2X HotStarTaq Plus Master Mix Kit (Qiagen, Venlo, Netherlands), 10–20 ng of genomic DNA and 0.2 μ M of each primer. Protocols for PCR reactions were taken from original sources (Table 1). Positive amplifications were confirmed with horizontal electrophoresis in agarose gels and PCR products were purified with AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA). A separate library was constructed for each primer set using NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs, Ipswich, MA, USA). All four libraries were sequenced using Illumina v2 Kit at a 2 \times 250 bp paired-end Illumina MiSeq platform. PCR products from each primer pair (88 individuals) were sequenced in a separate Illumina run. We also added 26 technical replicates across all four runs (26 different individuals genotyped twice using independent PCR amplifications).

Illumina data processing and allele validation

Illumina sequencing data were processed using the AmpliSequencing Analysis Tools (AmpliSAT) web server (Sebastian et al. 2016) and algorithms implemented therein. First, AmpliMERGE algorithm was used to merge paired-end Illumina reads (default settings). Second, we used the AmpliSAS tool to proceed with de-multiplexing, clustering, and filtering of merged sequences. The clustering stage uses the modified algorithm by Stutz and Bolnick (2014) to deal with genotyping errors associated with high-throughput sequencing techniques, i.e., to identify reads resulting from genotyping errors and to cluster them with reads identified as true alleles. Briefly, all variants in the amplicon are first ordered by depth and clustering starts from the most dominant sequence (highest depth). Each remaining amplicon sequence is compared with the dominant one and its sequencing/PCR errors (artefacts) are identified based on user-defined criteria by performing high-accuracy pairwise global alignments between the sequences (Sebastian et al. 2016). Here, we used default criteria for Illumina sequencing data (substitution errors—1%, indel errors—0.001%, minimum dominant frequency—25%), as recommended by Biedrzycka et al. (2017b). During the filtering stage, chimeras were detected using a set of specific rules (see Appendix 1 for details) and discarded. Sequences with the final (after clustering) read frequency of <3% per sample (amplicon) were also discarded (following AmpliSAT recommendations for Illumina sequencing). Minimum amplicon depth was 100 reads, while the maximum amplicon depth was set to 5000 reads (because of performance reasons). The AmpliSAS output for all four Illumina runs is shown in Appendix 2.

Although our genotyping approach targeted only fragments of MHC class I and class II molecules and these fragments originated from an unknown number of gene copies in an unknown synteny in the genome, we henceforth refer to these fragments as alleles. Different pairs of primers amplified sequences of different length (Table S2 and Fig. S1 in Appendix 1). At MHC class I, primers PP1_1

amplified a fragment of exon 3 (264 bp out of 276 bp total exon length, codon positions 5–92 within the exon) and primers PP1_2 amplified shorter fragment of the same exon (240 bp, codon positions 5–84) (Table S2 in Appendix 1). At MHC class II, primers PP2_1 amplified a fragment of exon 2 (219 bp out of 270 bp total exon length, codon positions 18–90 within the exon) along with four codons of intron 2 that were subsequently removed from the alignment (Table S2 in Appendix 1). Primers PP2_2 amplified shorter fragments of MHC class II exon 2 (168 bp, codon positions 18–73; Table S2 in Appendix 1). Longer sequences (obtained with PP1_1 and PP2_1 primers) were used for the analyses of polymorphism, recombination, selection, and phylogenetic clustering, as our priority was to obtain the results across the larger part of the exons rather than across the larger set of alleles. However, since none of our primer pairs was likely to amplify all existing alleles across all duplicated loci and null (unamplified) alleles were likely to differ between primer sets, we have repeated the analyses of selection using shorter sequences (PP1_2 and PP2_2) to test whether null alleles biased our results. Although our exon sequences were not complete, they covered the key fragments of the peptide-binding groove of MHC class I (ca. 95% of $\alpha 2$ domain) and class II (ca. 80% of $\beta 1$ domain) MHC molecules that are likely to be targeted by pathogen-driven selection in passerine birds (Minias et al. 2018). All alleles had cysteine amino acids at the highly conserved residues responsible for intra-domain disulfide bridge formation (positions 11 and 74 at MHC class I, position 75 at MHC class II; Fig. 3).

Copy number variation

Following a common practice in MHC research, the number of MHC gene copies in each species was estimated by dividing the maximum number of putatively functional MHC alleles detected per any individual within species by two (i.e., assuming heterozygosity at each locus) (reviewed in Minias et al. 2019). Here, we first assessed repeatability of copy number estimates between both independent pairs of primers by calculating intra-class correlation coefficients in the *irr* package (Gamer et al. 2012) developed for R statistical environment (R Foundation for Statistical Computing, Vienna, Austria) and, then, used the copy number estimates inferred from the combined data in all further analyses. Non-functional alleles (having stop codons or frame-shift mutations) were not included in copy number estimates. All the other alleles were recognized as putatively functional, although some of them could possibly contain non-functional mutations outside the regions that were targeted by genotyping. To reconstruct evolutionary history of copy number variation, we fitted six macroevolutionary models, which describe different evolutionary scenarios: (1)

Brownian Motion (BM)—constant rate neutral evolution (due to genetic drift) or fluctuating selection towards different evolutionary optima, (2) BM adjusted for λ —BM model with varying rates of genetic drift and fluctuating selection across the phylogeny (λ decreasing from 1 to 0 indicates increasing phylogenetic independence), (3) BM adjusted for δ —BM model with different rates of early (more ancestral) vs. late (more recent) evolution of the phylogenetic lineage, (4) time-dependent linear model—assumes linear changes in evolutionary rates, (5) time-dependent early burst model—assumes non-linearly decreasing evolutionary rates and (6) Ornstein–Uhlenbeck—stabilizing selection with a constant evolutionary optimum. All models were fitted to the data with the *fitContinuous* function in the *geiger* R package (Harmon et al. 2008) and the relative fit of all models was compared using the Akaike information criterion corrected for small sample sizes (ΔAIC_c). The rate of MHC copy number evolution was estimated using each of these models (1000 iterations per model). We also used the phylogenetic scaling parameter λ (BM model adjusted for λ ; Pagel 1999) to estimate phylogenetic signal, which describes statistical dependence in trait values (MHC gene copy numbers) among species due to their phylogenetic relationships (the tendency of related species to resemble each other) (Revell et al. 2008). In general, phylogenetic signal can vary between 0 and 1, where $\lambda = 0$ indicates complete phylogenetic independence of a trait, whereas $\lambda = 1$ indicates that trait evolution is completely determined phylogenetically (Freckleton et al. 2002). The ancestral states of MHC copy numbers at the root of the phylogeny (trait value expected in the last common ancestor of our study species) were reconstructed using the maximum-likelihood approach (under the assumptions of BM process), as implemented in *fastAnc* function in the *phytools* R package (Revell 2012). Phylogenetic relationships between species were based on a complete avian time-calibrated phylogeny developed by Jetz et al. (2012) with a backbone topology from Ericson et al. (2006). To account for phylogenetic uncertainty, we generated a consensus phylogeny from a random sample of one thousand trees downloaded from the BirdTree web server (Jetz et al. 2012). Differences in the number of MHC class I and class II copies detected per species were tested with *t*-test for dependent samples, while the linear association between MHC class I and class II copy numbers across species was tested with Pearson correlation coefficient.

Sequence polymorphism

We used DnaSP v.6.10.3 software (Rozas et al. 2017) to calculate the number of segregating sites, total number of mutations and nucleotide diversity, as the basic measure of

sequences polymorphism. We also used *DistCalc* function from the *MHCtools* R package (Roved 2019) to calculate Grantham (1974) and Sandberg et al. (1998) distances between amino acid sequences, as they take physiochemical properties of amino acids into account. All polymorphism measures were calculated separately for each clade, as well as across all clades. Since we had no reliable information on the breeding origin of sampled birds (most individuals were captured during migration), we did not compute Tajima's *D* values.

Recombination

To quantify recombination signal at the MHC class I and class II, we used RDP v.4.97 software (Martin et al. 2015), which implements several different algorithms developed to detect recombinant sequences. The following approaches were used to assess recombination in our data: Maxchi (Smith 1992), BootScan (Salminen et al. 1995), Genconv (Padidam et al. 1999), SiScan (Gibbs et al. 2000), RDP (Martin and Rybicki 2000), Chimaera (Posada and Crandall 2001) and 3Seq (Boni et al. 2007). All analyses were run using default settings, statistical significance threshold of $P = 0.05$, and Bonferroni correction for multiple comparisons. A recombination event was recognized when supported by two or more algorithms, all events recognized by a single algorithm were discarded. Different recombination events were identified based on the different location of recombination breakpoints along exon sequences. Since MHC polymorphism may be retained beyond species divergence, recombination signal can be traced over long evolutionary times (Minias et al. 2016) and, thus, we conducted the analyses across all species from all clades. Recombination signal was also assessed as the number of breakpoints within 100 nucleotide window and the presence of recombination hot and cold spots was tested with the local hot/cold-spot test (1000 permutations), as implemented in RDP software.

Selection

Patterns of selection were assessed with the relative rate of nonsynonymous (amino acid altering) nucleotide substitutions per nonsynonymous site to synonymous (silent) nucleotide substitutions per synonymous site (dN/dS). Nonsynonymous mutations are expected to accumulate faster than synonymous mutations under positive (diversifying) selection, resulting in $dN/dS > 1$. In contrast, negative (purifying) selection removes nonsynonymous mutations and, thus, synonymous mutations accumulate at a faster rate ($dN/dS < 1$). Finally, lack of significant deviations from $dN/dS = 1$ indicates no signature of selection, consistent with neutral evolution. Following recommendations by Kryazhimskiy and Plotkin (2008), substitution rates were estimated at the inter-specific levels, both

within and across all three phylogenetic clades. Since selection estimates can be biased by the presence of recombinant sequences, which affect tree topologies used to infer substitution rates (Anisimova et al. 2003), the dN/dS ratios were computed using non-recombinant sequences only, except for estimates across all clades, which were computed both with and without recombinant sequences. Codon-specific signatures of pervasive (constant along the entire phylogeny) positive and negative selection were inferred using Bayesian (Fast Unconstrained Bayesian Approximation, FUBAR) and maximum-likelihood (fixed effect likelihood, FEL) approaches (Kosakovsky Pond and Frost 2005; Murrell et al. 2013). Also, episodic (under a proportion of branches) positive selection was assessed using Mixed Effects Model of Evolution (MEME) (Murrell et al. 2012). All analyses were run at the Datamonkey web server (Weaver et al. 2018) under default settings and input trees inferred from the sequence alignments. Residues with posterior probabilities > 0.95 (FUBAR) or statistical significance $P < 0.05$ (FEL, MEME) were considered to have enough support for selection signal. As the exact structure of MHC molecules, including the location of the peptide-binding groove, has not yet been resolved in our study species (and generally in birds), the dN/dS ratios were calculated for different sets of MHC class I exon 3 and class II exon 2 residues: (1) all residues; (2) 20 most positively selected residues; (3) putative peptide-binding residues of passerine birds, as inferred from the global analysis of selection at the avian MHC (Minias et al. 2018); and (4) human peptide-binding residues, as inferred from the crystallographic structure of MHC molecules (class I: Saper et al. 1991; class II: Brown et al. 1993).

Phylogenetic clustering of alleles

Phylogenetic relationships between MHC alleles were inferred using two approaches. First, we used the approximately maximum-likelihood approach implemented in FastTree v.2.1.5 software (Price et al. 2010), where initial topologies were inferred using a neighbour-joining approach and refined with subtree-pruning-regrafting, as well as minimum-evolution and maximum-likelihood nearest-neighbour interchanges. Local node support values were estimated based on Shimodaira–Hasegawa (1999) tests. Second, we used Bayesian inference from MrBayes v.3.2.6 software (Ronquist et al. 2012), which implements Markov chain Monte Carlo (MCMC) algorithms to approximate the posterior probabilities of trees (MCMC settings: 4 chains, 120,000 chain length, 20,000 burn-in period and 500 sampling frequency). While neither of these approaches is valid for ancestral state reconstruction, they are expected to provide reliable information on allele clustering patterns, although we acknowledge that the results may be affected by recombination (gene conversion)

mechanism typical for the MHC. Phylogenetic analyses were run separately for MHC class I and class II alleles. In all analyses, we used a general time-reversible model of nucleotide substitutions with a discrete Gamma distribution to account for varying rates of evolution at different sites, but for MHC class I we also set the proportion of invariable sites (as determined with model selection analysis in MEGA v6.06; Tamura et al. 2007). MHC sequences from the lesser kestrel *Falco naumanni* (GenBank Nos: EU120676 for class I, EU107738 for class II) were used as outgroups, as Falconiformes form a closely related ancestral clade for Passeriformes (Jetz et al. 2012).

Results

Illumina sequencing with different pairs of primers

After processing, the total depth of alleles varied between 2083 and 4142 reads per sample (amplicon), depending on the pair of primers (Table S2 in Appendix 1). We retrieved 439 (PP1_1) and 387 (PP1_2) MHC class I alleles, while at the MHC class II we retrieved 306 (PP2_1) and 372 (PP2_2) alleles. Reproducibility of allele identification was 92%, as assessed with technical replicates. No evidence was found for a correlation between read number and number of alleles detected for any of the runs (Pearson correlation coefficients, all $P > 0.05$). The number of non-functional alleles was relatively low (2.3–2.7% for class I; 0.8–4.2% for class II). The highest number of non-functional alleles was recorded in Fringillinae (five MHC class I and 11 MHC class II sequences). Much lower numbers (1–2) of non-functional alleles were recorded in several other species (*A. flammea*, *C. carduelis*, and *P. pyrrhula* at class I; *C. coccothraustes*, *E. schoeniclus* at class II), while in the remaining species no non-functional alleles were recorded. All non-functional alleles were discarded, resulting in 427 (PP1_1) vs. 377 (PP1_2) functional MHC class I alleles (337 overlapping between both primer pairs), and 289 (PP2_1) vs. 364 (PP2_2) functional MHC alleles (210 overlapping between both primer pairs).

The number of alleles detected per individual was much more repeatable between the pairs of primers at class I ($R = 0.75$, 95% CI: 0.65–0.83) than class II ($R = 0.35$, 95% CI: 0.13–0.53). A similar pattern was found for the number of gene copies estimated for each species based on the maximum number of putatively functional MHC alleles detected per individual (class I: $R = 0.83$, 95% CI: 0.46–0.95; class II: $R = 0.57$, 95% CI: –0.05 to 0.87). Higher repeatability for class I could be enhanced by relatively similar forward primers (88.2% vs. 40.0% sequence similarity for MHC class I and class II, respectively; Table 1). At class I, the average number of alleles detected per individual was

7.67 ± 0.43 for PP1_1 and 7.26 ± 0.30 for PP1_2, while data combined for both primer pairs gave 8.88 ± 0.43 alleles retrieved per individual. Large differences in the number of alleles per individual between primer pairs were detected in five species, where PP1_1 retrieved more alleles in *E. citrinella* (on average 2.63 ± 0.92 more alleles), *F. coelebs* (2.88 ± 0.55 alleles) and *F. montifringilla* (3.50 ± 0.42 alleles), while PP1_2 retrieved more alleles in *E. schoeniclus* (3.25 ± 0.62 alleles) and *C. coccothraustes* (3.13 ± 0.55 alleles) (Fig. S2A in Appendix 1). At the class II, the average number of alleles detected per individual was 5.05 ± 0.24 (PP2_1) and 6.31 ± 0.23 (PP2_2), but combination of data for both primer pairs resulted in an average of 8.33 ± 0.31 alleles per individual. The largest differences were recorded in Fringillinae, where PP2_2 retrieved on average 2.25 ± 0.96 (*F. coelebs*) and 4.75 ± 0.70 (*F. montifringilla*) more alleles per individual than PP2_1, whereas PP2_1 retrieved 2.00 ± 0.93 more alleles per individual in *P. pyrrhula* (Fig. S2B in Appendix 1). Any single primer pair underestimated the number of functional gene copies per species by an average of 0.86 ± 0.24 (MHC class I) and 1.45 ± 0.23 (MHC class II), being consistent with an overall underestimation rate of 20.2% when compared to data combined across both pairs of primers.

Copy number variation

The total number of alleles retrieved within individuals (across both pairs of primers) was 2–20 (class I) and 4–18 (class II). The number of gene copies in each species ranged from two (*P. pyrrhula*) to ten (*F. coelebs*) at the MHC class I, and from three (*S. spinus*) to nine (*E. citrinella*) at the MHC class II (Fig. 1). Phylogenetic signal was much lower at MHC class I than class II (0.19 vs. 0.64). The evolution of copy number at both MHC classes was best explained by the BM model (Table 2). Ancestral states at the root of the phylogeny were estimated at intermediate values (5.75 class I copies and 6.03 class II copies according to the best fitting model, Table 2), suggesting that some taxa lost, while the other gained MHC copies during the evolution of this passerine lineage. Ancestral states at the root node for each of three clades were highest for Fringillinae (8.42 class I copies and 7.22 class II copies) and lowest for Carduelinae (4.99 class I copies and 5.32 class II copies). Evolutionary rate was higher for MHC class I rather than class II copy number evolution (Table 2).

Sequence polymorphism

Although MHC class I had higher allelic diversity than class II (427 vs. 289 alleles, as retrieved across all species with PP1_1 and PP2_1 primer pairs), sequence polymorphism at the nucleotide level was greater at the MHC class II. MHC

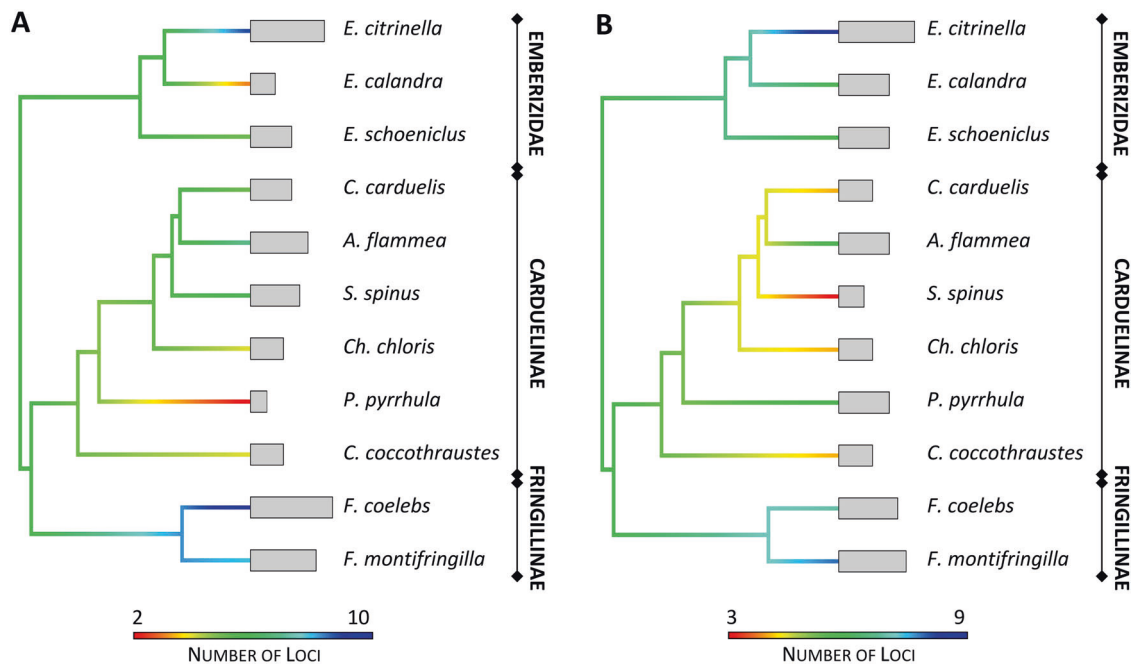


Fig. 1 MHC gene copy numbers. Ancestral character estimation of copy numbers at MHC class I (A) and class II (B) along the branches and nodes of the phylogeny in finches and buntings. Bars associated with each terminal node indicate raw number of gene copies (loci).

Table 2 Relative fit (ΔAIC_C) and parameter estimates of models describing the evolution of gene copy number at MHC class I and class II in finches and buntings.

MHC	Evolution model	AIC_C	ΔAIC_C	Parameters		
				Evolutionary rate (σ^2)	Ancestral state (z_0)	Model-specific parameters
Class I	Brownian Motion	57.15	0.00	0.33	5.75	–
	Ornstein–Uhlenbeck	59.89	2.74	1.11	5.66	$\alpha = 0.09$
	Brownian Motion adjusted for δ	59.92	2.77	0.22	5.59	$\delta = 2.99^a$
	Brownian Motion adjusted for λ	60.01	2.86	0.21	5.78	$\lambda = 0.19$
	Time dependent (linear)	60.13	2.98	<0.001	5.62	$b = 99.9^a$
	Time dependent (early burst)	72.24	15.09	0.10	5.75	$b = 0^a$
Class II	Brownian Motion	47.99	0.00	0.16	6.03	–
	Brownian Motion adjusted for λ	51.24	3.25	0.11	5.99	$\lambda = 0.64$
	Ornstein–Uhlenbeck	51.32	3.33	0.24	5.92	$\alpha = 0.03$
	Brownian Motion adjusted for δ	51.35	3.36	0.11	5.89	$\delta = 1.90$
	Time dependent (early burst)	51.92	3.93	0.14	6.03	$b = 0^a$
	Time dependent (linear)	52.18	4.19	0.008	5.89	$b = 0.86$

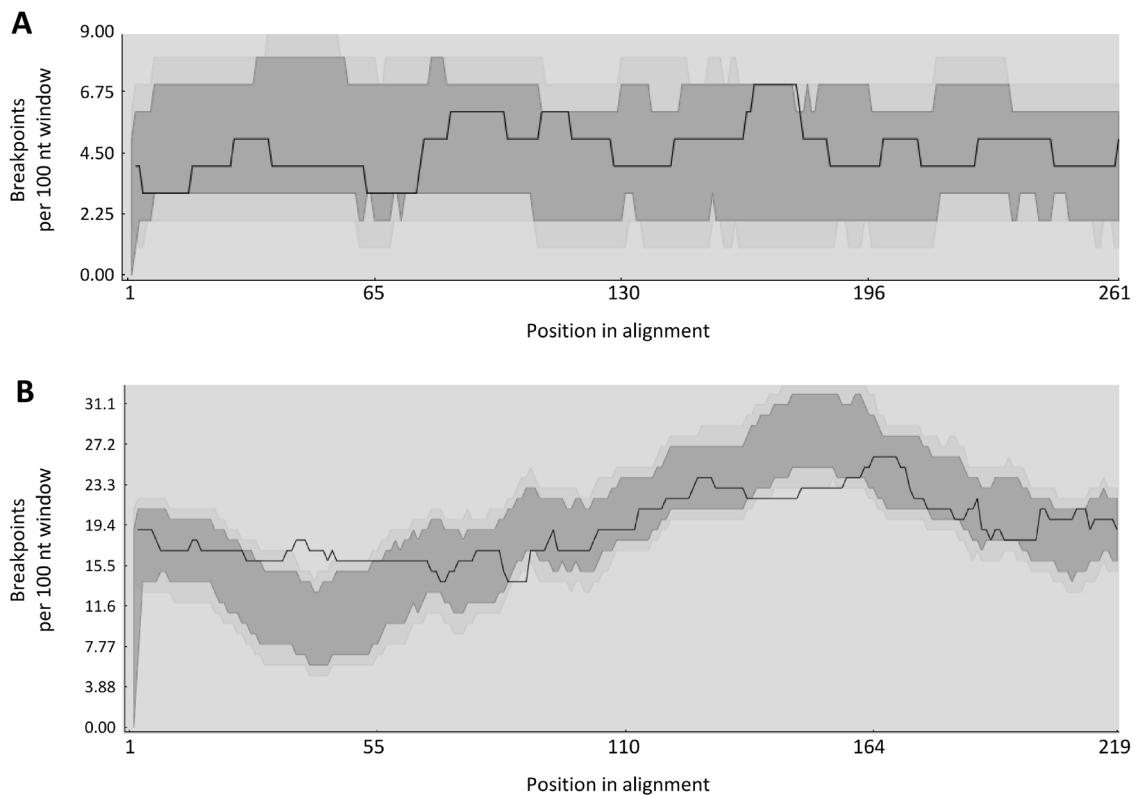
Parameters that appeared at default bounds after 1000 iterations are marked with letter a.

class II exon 2 had higher proportion of segregating sites (84.5%) and more mutations ($n = 364$) than MHC class I exon 3 (74.3% segregating sites; 313 mutations) (Table 3). Nucleotide diversity and amino acid distances (Grantham

and Sandberg) were also higher at MHC class II than class I (Table 3). This pattern was primarily apparent in Emberizidae and Carduelinae, but not in Fringillinae (Table 3). Indel mutations were recorded within exons of both MHC

Table 3 Polymorphism of MHC class I exon 3 and class II exon 2 within and across three passerine clades of finches and buntings ($n = 8$ individuals per species).

MHC	Exon	Exon length	Passerine clade	No. species	No. alleles	No. segregating sites	No. mutations	Nucleotide diversity	Grantham distance (mean \pm SE)	Sandberg distance (mean \pm SE)
Class I	3	264 bp	Emberizidae	3	91	128	163	0.100	14.08 \pm 0.12	0.980 \pm 0.008
			Carduelinae	6	185	149	207	0.123	16.54 \pm 0.06	1.082 \pm 0.004
			Fringillinae	2	151	144	194	0.120	14.49 \pm 0.05	0.996 \pm 0.003
			ALL	11	427	194	313	0.138	17.72 \pm 0.02	1.196 \pm 0.001
Class II	2	222 bp	Emberizidae	3	91	150	228	0.190	25.27 \pm 0.17	1.738 \pm 0.012
			Carduelinae	6	147	152	262	0.181	23.62 \pm 0.08	1.609 \pm 0.005
			Fringillinae	2	52	103	113	0.067	10.66 \pm 0.24	0.741 \pm 0.016
			ALL	11	293	185	364	0.219	28.81 \pm 0.05	1.969 \pm 0.003

**Fig. 2 Recombination at the MHC.** Recombination signal at the MHC class I exon 3 (**A**) and class II exon 2 (**B**), as assessed with the local hot/cold-spot test. Positions in alignment where black line emerges above the grey areas indicate a recombination hot spot, while positions where it drops below the grey areas indicate a recombination cold spot.

the absence of recombination hot and cold spots (as assessed with the local hot/cold-spot test). Positions in alignment where black line emerges above the grey areas indicate a recombination hot spot, while positions where it drops below the grey areas indicate a recombination cold spot.

classes. A single-codon deletion at position 61 of the MHC class I exon 3 was recorded in 54.9% sequences across all species. Similarly, a single-codon deletion was recorded at position 74 of the MHC class II exon 2, but only in Emberizidae (*E. citrinella* and *E. schoeniclus*; 1.4% of all sequences). Also, a single-codon insertion after position 85 was recorded within the same exon in Carduelinae (*A. flammea* and *Ch. chloris*; 7.5% of all sequences).

Recombination

Recombination signal was much stronger at the MHC class II when compared with class I. We identified 20 recombination events at the class II, while only six recombination events were identified at the class I. The number of breakpoints per 100 nt window was estimated 14–26 for the MHC class II and 3–7 for MHC class I (Fig. 2). Proportion

of recombinant sequences was also significantly higher for the MHC class II than class I (27.7 vs. 8.2%; $G = 23.97$, $df = 1$, $P < 0.001$). The average number of recombinant sequences per recombination event was similar between the MHC classes (5.83 ± 2.68 vs. 4.00 ± 1.10 ; $t = 0.74$, $df = 24$, $P = 0.47$). The most recombination events were recorded within Carduelinae (66.7% at class I and 90% at class II), while MHC sequences from other clades showed much lower evidence for recombination, but this may be due to differences in the number of species genotyped in each clade.

Selection

Signature of positive selection was consistently stronger at the MHC class I than class II, both within and across all three clades. The number of residues under pervasive positive selection in different clades was 10–17 or 11–16 at class I (11.4%–19.3% of all residues), as recognized with Bayesian (FUBAR) or maximum-likelihood (FEL) algorithms, respectively (Table 4). At class II, the number of residues under pervasive positive selection was estimated at 2–8 (FUBAR) or 1–10 (FEL) (1.4–13.7% of all residues). Across all clades, the number of residues under pervasive selection was estimated at 18 (20.5%) and 9–10 (12.3%–13.7%) at class I and class II, respectively (Table 4 and Fig. 3). The number of residues under episodic positive selection was higher at the MHC class I in two clades: Emberizidae (20 vs. 12; 22.7 vs. 16.4% of all residues) and

Fringillinae (20 vs. 2; 22.7 vs. 2.7% of all residues), while contrasting pattern was found for Carduelinae (15 vs. 21 residues at class I and class II, respectively; 17.0 vs. 28.8% of all residues) (Table 4). The dN/dS ratios at the most positively selected residues and at the passerine PBR were consistently higher at the MHC class I than class II, both within and across clades (Table 4), providing further support for stronger signature of positive selection at the MHC class I. Identification of positively selected residues (across clades) was highly repeatable between FUBAR and FEL algorithm (class I: $R = 0.93$, 95% CI: 0.89–0.95; MHC class II: $R = 0.88$, 95% CI: 0.81–0.92). Positions of positively selected residues identified across all three clades were moderately repeatable with positions of the passerine PBR (class I: $R = 0.47$, 95% CI: 0.28–0.61; MHC class II: $R = 0.51$, 95% CI: 0.32–0.66) and poorly repeatable with the human PBR (class I: $R = 0.20$, 95% CI: –0.02 to 0.39; MHC class II: $R = 0.09$, 95% CI: –0.13 to 0.30) (Fig. 3). We also found strong repeatability in selection signal ($0.60 < R < 0.97$, all $P < 0.001$; Table S3) between two independent sets of primers (PP1_1 and PP1_2 for MHC class I; PP2_1 and PP_2 for MHC class II), which indicated that null alleles were unlikely to introduce significant biases.

Phylogenetic clustering

Neither MHC class I nor II alleles clustered by family, subfamily (clade) or species, although phylogenetic pattern

Table 4 Selection signature at the MHC class I exon 3 (264 bp fragment) and class II exon 2 (222 bp fragment) within and across three passerine clades of finches and buntings.

MHC	Passerine clade	Number of residues under positive selection			dN/dS			
		Pervasive		Episodic	All residues	Positively selected residues	Passerine PBR	Human PBR
		FUBAR	FEL					
Class I	Emberizidae	11	14	20	0.82	5.02	2.37	1.14
	Carduelinae	10	11	15	0.68	3.45	2.12	1.38
	Fringillinae	17	16	20	0.74	4.46	3.04	1.88
	ALL (no recomb.)	17	17	25	0.79	4.58	2.64	1.29
	ALL (recomb.)	18	18	26	0.73	4.25	2.44	1.34
Class II	Emberizidae	8	6	12	0.77	2.27	1.49	1.53
	Carduelinae	7	10	21	0.78	3.05	2.08	1.17
	Fringillinae	2	1	2	0.30	2.37	0.98	0.27
	ALL (no recomb.)	8	8	27	0.79	2.62	1.86	1.34
	ALL (recomb.)	10	9	30	0.91	2.57	2.11	1.33

Strength of selection was measured with the number of residues (sites) under pervasive and episodic positive selection (as inferred with FUBAR/FEL and MEME algorithms, respectively) and with the relative rates of nonsynonymous to synonymous substitutions (dN/dS) at: (1) all residues; (2) 20 most positively selected residues; (3) putative peptide-binding region (PBR) of passerines, as inferred from the global analysis of selection at the avian MHC (Minias et al. 2018); and (4) human peptide-binding residues, as inferred from the crystallographic structure of MHC molecules (Saper et al. 1991; Brown et al. 1993). All estimates were inferred for non-recombinant sequences only, except for estimates across all clades, which were computed for data with (recomb.) and without (no recomb.) recombinant sequences.

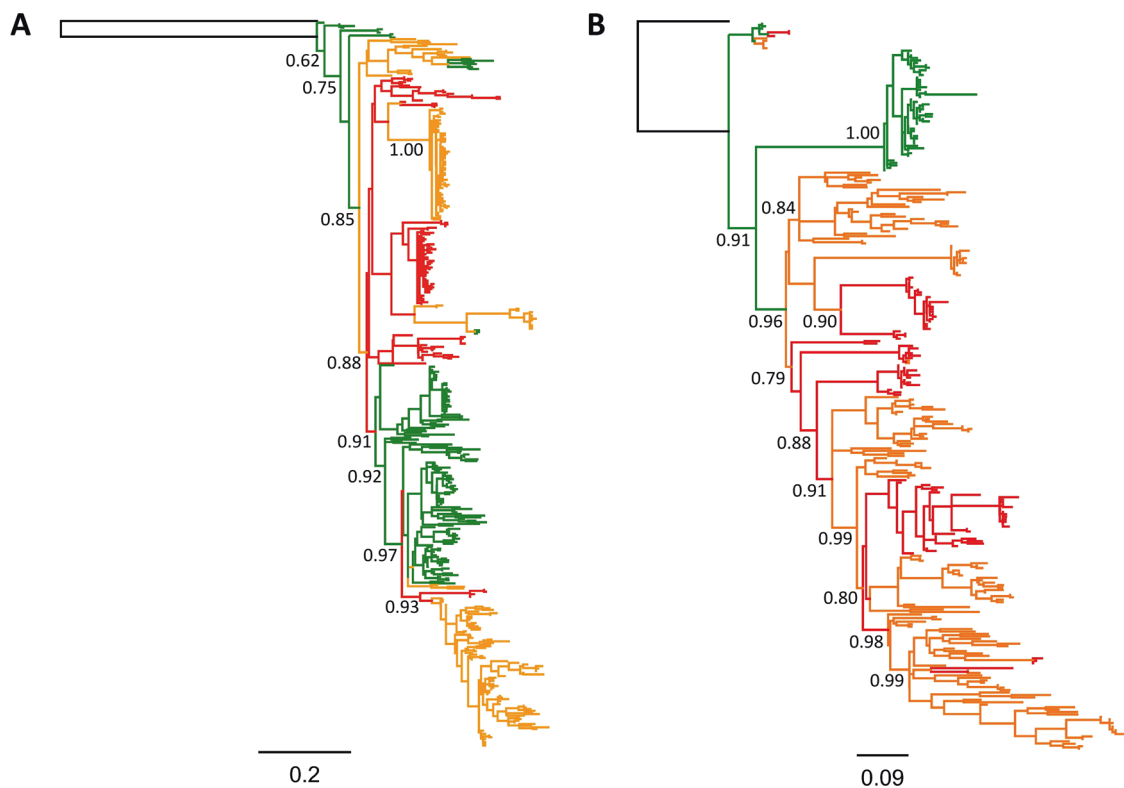


Fig. 4 Phylogenetic clustering of MHC alleles. Consensus maximum-likelihood topology for MHC class I exon 3 (**A**) and MHC class II exon 2 (**B**) in Emberizidae (red), Carduelinae (orange), and Fringillinae (green). Lesser kestrel was used as an outgroup to root the

trees (black). Local bootstrap support is provided for key nodes (only values >0.5 presented). Scale bar indicates genetic distance (nucleotide substitutions per site).

positive selection than MHC class II and this might have been a key evolutionary force that generated and maintained higher allele diversity (allele numbers) at the MHC class I in our study species. Despite weaker selection and lower allele diversity, MHC class II genes showed greater allele divergence (in terms of nucleotide diversity) and had much greater recombination (gene conversion) signal. Contrasting selection regimes did not generate any apparent variation in copy numbers between MHC class I and class II, which both evolved via fluctuating selection and drift (consistent with BM evolution; Revell et al. 2008). However, the rate of copy number evolution was higher at the MHC class I than class II, resulting in higher variation of MHC class I copy numbers among closely related species (lower phylogenetic signal).

Copy number variation

The first aim of our study was to reconstruct the copy number evolution at the MHC in finches and buntings. The highest numbers of gene copies (≥ 8 class I and ≥ 7 class II) was recorded in *E. citrinella* (Emberizidae) and the two *Fringilla* species (Fringillinae), suggesting that duplication rate in these taxa could have been enhanced by elevated

selective pressure from both intra-cellular and extra-cellular pathogens. Both MHC class I and class II were found to have evolved consistently with the BM model, indicating a key role of fluctuating selection in the copy number evolution. Previous analyses in passerines suggested that MHC class I copy numbers might have evolved consistently with a BM model (O'Connor et al. 2016), but similar analyses of passerine MHC class II were lacking. On the other hand, a large-scale comparative analysis of MHC genes across the entire avian tree (passerines and non-passerines combined) provided support for highly contrasting patterns of copy number evolution between MHC classes I and II (Minias et al. 2019). The results of this study suggest that broad generalizations may not well reflect the patterns at the finer phylogenetic scale, as we showed that similar mechanisms were responsible for the evolution of gene copy numbers at both MHC classes I and II in our study system (Fringillidae and Emberizidae passerines). Despite this similarity, evolutionary rate was higher at class I than class II and there was a greater variation in MHC class I copy numbers among closely related species (lower phylogenetic signal), which may suggest that selective regimes at class I are more evolutionarily labile and species specific. To quantify the number of MHC gene copies, we followed an approach by

O'Connor et al. (2016), who used two independent pairs of primers to determine the total number of alleles recorded in each individual. Here, we confirmed that this approach can substantially increase reliability of copy number estimates, as using a single pair of primers would underestimate copy numbers by, on average, ca. 15% at class I and 25% at class II across all species, because of allele-specific amplifications by each primer pair (27.8% MHC class I and 52.5% MHC class II alleles were amplified in a primer-specific manner). Taking this into account, we highly recommend using multiple pairs of primers in future research on copy number variation at the MHC in non-model vertebrates species.

Selection patterns

Stronger signature of positive selection at the MHC class I (when compared with class II) is consistent with expectations from the recent global analysis of selection at the avian MHC (Minias et al. 2018). Although Minias et al. (2018) recorded higher number of positively selected sites and higher nucleotide substitution rates (dN/dS ratios) at the class I exon 3 than MHC class II exon 2 across over 20 families of passerine birds, our study species of Fringillidae and Emberizidae were virtually unrepresented in these analyses (only two MHC class II alleles from two species included). More importantly, previous analyses were based on highly unbalanced GenBank data (in terms of sample sizes and species composition) that were generated using different methodological approaches, which could possibly introduce biases in the comparisons between the two MHC classes. Here, we attempted a rigorous comparison of selection patterns between MHC class I and class II across multiple passerine species, based on uniform methodology and balanced sample sizes. The results of our selection analysis in finches and buntings revealed stronger selection at MHC class I than class II, but the magnitude of these differences largely differed between clades. The most apparent differences in pervasive positive selection were recorded in Fringillinae (16–17 vs. 1–2 positively selected residues at MHC classes I and II, respectively), while Carduelinae showed weakest differences (10–11 vs. 7–10 positively selected residues at MHC classes I and II, respectively). The analysis of dN/dS ratio provided a very similar picture, suggesting that selective regimes at the two MHC classes may be specific to particular avian lineages. So far, comparative analyses of selection patterns at the MHC class I and class II in non-model passerine birds have been scarce and the only exception that we are aware of is the domestic sparrow *Passer domesticus*, in which stronger signature of positive selection was found at the MHC class II than at class I (Borg et al. 2011), being in stark contrast to our findings for finches and buntings. MHC research in non-

passerine birds from various orders also seems to provide clear support for stronger positive selection at the MHC class II. For example, a higher proportion of positively selected residues was recorded at MHC class II exon 2 (13%) than MHC class I exon 3 (7%) across six flamingo species (order Phoenicopteriformes) (Gillingham et al. 2016). Even greater differences were recorded in five species of prairie grouse (order Galliformes) with 3.4% and 17.7% of residues being under positive selection at the MHC class I and class II, respectively (Minias et al. 2016). This corresponded with ca. threefold difference in the dN/dS ratio, being higher at the MHC class II (Minias et al. 2016). Variation in the strength of selection signal between MHC class I and class II is traditionally attributed to differences in selective pressures from intra-cellular and extra-cellular pathogens, and although this may be speculative, our findings seem to suggest that intra-cellular pathogens could possibly be considered as the major selective agent for finches and buntings. In general, all our study species are small (mean body mass of 27.6 ± 4.3 g) and are expected to harbour less diverse faunas of extra-cellular parasites than much larger non-passerine species (e.g., Morand and Poulin 1998), which may explain greater evolutionary importance of MHC class I in these taxa.

Polymorphism and phylogenetic clustering

As expected, stronger positive selection at the MHC class I resulted in higher allele diversity at these genes (427 vs. 293 alleles at classes I and II, respectively). However, although class II alleles were less diverse in terms of their numbers, they showed greater divergence at the nucleotide level, with almost 60% higher nucleotide diversity compared to the MHC class I alleles (0.219 vs. 0.138 across all species). This pattern may indicate that duplications of MHC class II genes occurred earlier in the evolutionary history of these lineages than duplications at the MHC class I, consistently with lower evolutionary rate and higher phylogenetic signal in MHC class II copy numbers, and thus alleles from different class II loci might have had longer time to accumulate independent mutations and diverge. This scenario seems to be supported by phylogenetic inference, providing evidence for the presence of a small MHC class II cluster, where alleles from all three subfamilies clustered together. This could possibly represent an old evolutionary lineage of MHC class II alleles that was retained beyond species divergence. At the same time, all the other class II alleles of Fringillinae formed a distinct cluster that was well separated from both Carduelinae and Emberizidae sequences, also suggesting that MHC class II genes may show low level of orthology between these groups. This phylogenetic pattern was clearly inconsistent with the evolutionary history of our study taxa, where the Emberizidae lineage first separated

from the *Carduelinae*–*Fringillinae* lineage (ca. 26 million years ago; Jetz et al. 2012), and thus was expected to show strongest genetic divergence. Evolutionary relationships between taxa were even less apparent at the MHC class I, where alleles from all clades were highly interspersed across the entire tree, providing evidence for strong trans-species polymorphism. The presence of trans-species polymorphism was also supported by several cases of allele sharing between species, which were more frequent at MHC class I than class II. The pattern of trans-species polymorphism and allele sharing among species at the MHC was previously described in other avian lineages (Kikkawa et al. 2009; Eimes et al. 2015) and other vertebrates (e.g., Cutrera and Lacey 2007), and it is hypothesized to occur under strong balancing selection, which may enhance maintaining favourable alleles through long evolutionary times (Takahata 1990).

Recombination signal

Higher nucleotide divergence of MHC class II alleles was accompanied with higher recombination (gene conversion) rate at these genes. This pattern was rather unexpected, as gene conversion does not generate *de novo* variation at the MHC (and thus it may not necessarily contribute to sequence divergence), but it generates new haplotypes by shuffling existing variation within and between loci (and thus it should contribute to allele richness; Högstrand and Böhme 1999). In fact, gene conversion is recognized as one of the primary mechanisms that maintain allele diversity at the MHC and it is known to be particularly important for regenerating MHC haplotype variation in bottlenecked populations, as the gene conversion rate may greatly exceed the rate of point mutations (Spurgin et al. 2011). Surprisingly, the recombination rate in our study system (higher at class II) followed a pattern that clearly contrasted with allele diversity (higher at class I). As much as 27.7% of MHC class II alleles were identified to show a signal of recombination (compared to 8.2% at class I) and we detected 14–26 recombination breakpoints per 100 nt at the MHC class II (compared to 3–7 breakpoints at class I). Since we could not assign alleles to loci, it is impossible to conclude whether this higher recombination rate at the MHC class II was primarily driven by within- or between-locus DNA transfer. Higher between-locus gene conversion may lead to the concerted evolution of MHC loci, where MHC sequences derived from various loci become homologous and their orthology can be masked (Witzell et al. 1999). In our study, MHC class II alleles clustered more by taxon (family/subfamily) when compared with MHC class I, which may suggest a higher rate of concerted evolution at MHC class II than class I in finches and buntings. In birds, concerted evolution has usually been detected at MHC class

II (e.g., Miller and Lambert 2004; Li et al. 2011; Gillingham et al. 2016; Goebel et al. 2017), but this was primarily attributed to scarcity of similar research on MHC class I (Promerová et al. 2009). Nevertheless, taxon-specific analyses of scarlet rosefinches *Carpodacus erythrinus* and blue tits *Cyanistes caeruleus* revealed patterns consistent with concerted evolution at the MHC class I (Promerová et al. 2009; Wutzler et al. 2012). Our direct comparison of MHC classes I and II suggests that stronger concerted evolution at class II may reflect different evolutionary trajectories of the two classes rather than being a mere artefact resulting from publication bias.

Methodological limitations

Rapid development of next-generation sequencing technology within the last two decades sparked an unprecedented advancement in research on multi-gene families, such as the MHC. Although short-read sequencing can provide an easy way to obtain extensive information on relatively short fragments of targeted genes in a variety of non-model species, this approach certainly suffers from major drawbacks, as it does not provide a holistic view on the molecular evolution of entire genes and usually cannot assign allele fragments to particular loci. Previous research using longer sequences (including intron regions) or implementing locus-specific primers provided novel and invaluable insights into gene orthology and long-term evolutionary dynamics of the avian MHC (e.g., Burri et al. 2008, 2010; Cloutier et al. 2011). Third-generation sequencing, which allows to reliably assemble long regions with repetitive elements, promises the next important step in MHC research, providing information on the arrangement of the MHC genes at the genomic scale (O'Connor et al. 2019; He et al. 2021). However, despite its inherent limitations, we believe that research based on short-read sequencing can still provide a valuable contribution to our understanding of the MHC evolution when used in a phylogenetically robust (multi-species) context, which is often unattainable by other, much more laborious and high-cost approaches.

Conclusions

In conclusion, our study constitutes one of few existing evolutionary comparisons of avian MHC class I and class II genes in a multi-species approach. So far, research of this kind primarily focused on non-passerine avian lineages, which generally have a different architecture of MHC region to passerines (e.g., much lower copy numbers, Minias et al. 2019). Our molecular analyses of the MHC in two diverse families of Old World passerines, finches and

buntings, provide support for contrasting evolutionary trajectories of class I and class II genes, where stronger positive selection and faster evolutionary rate of gene copy numbers were recorded at MHC class I, whereas MHC class II showed a stronger gene conversion (recombination) signal, greater sequence polymorphism and stronger signal of concerted evolution. While our results suggest that intra-cellular and extra-cellular pathogens can exert different evolutionary effects on their hosts, at the same time they reveal fine-scale (between families and subfamilies) variation in the mechanisms that drive the evolution of passerine MHC classes I and II. We strongly recommend re-focusing MHC research from single-species and single-class approaches towards multi-species analyses of both MHC classes, which may substantially increase our understanding of evolution of these key immune genes in different phylogenetic vertebrate lineages.

Data availability

All sequences generated and used in this study have been deposited in GenBank (Nos: MW740435–MW741429).

Acknowledgements We thank all participants of bird ringing projects (Carpatica, Jeziorsko and Operation Baltic) that contributed to sample collection. At two coastal Operation Baltic ringing stations, Mierzeja Wiślana at Vistula Spit, and Bukowo-Kopań, MR, Krzysztof Stepniewski and Michał Redliski collected samples and collated bird data. Fieldwork at Operation Baltic was supported by Special Research Facility grants (SPUB) of the Polish Ministry of Science and Higher Education (203733/E-335/SPUB/2016/4) to the Bird Migration Research Station, University of Gdańsk. IC was supported by the Institute of Biology Bucharest, Romanian Academy (RO1567-IBB04/2020). The study was financially supported by the research grant of the National Science Centre in Poland (2015/19/D/NZ8/01310). We thank three anonymous reviewers and the Editor, Barbara Mable, for constructive comments on the earlier draft of the paper.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethical approval Catching of birds was authorized by the Polish Academy of Sciences, with the approval of the General Directorate for Environmental Protection, Poland. Field research at Bukowo and Mierzeja Wiślana was approved by the Marine Office in Słupsk and Gdańsk, respectively. The study was performed under the permit from the Local Bioethical Committee for Experiments on Animals in Łódź and appropriate Regional Environmental Protection Directorates.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Aguilar A, Edwards SV, Smith TB, Wayne RK (2006) Patterns of variation in MHC class II β loci of the little greenbul

- (*Andropadus virens*) with comments on MHC evolution in birds. *J Hered* 97:133–142
- Alcaide M, Liu M, Edwards SV (2013) Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ* 1:e86
- Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164:1229–1236
- Arnaiz-Villena A, Lowy E, Ruiz-del-Valle V, Westerdahl H, Moscoso J, Serrano-Vela JI et al. (2007) Evolution of the major histocompatibility complex class I genes in *Serinus canaria* from the Canary Islands is different from that of Asian and African continental *Serinus* species. *J Ornithol* 148:479–484
- Balasubramaniam S, Bray RD, Mulder RA, Sunnucks P, Pavlova A, Melville J (2016) New data from basal Australian songbird lineages show that complex structure of MHC class II β genes has early evolutionary origins within passerines. *BMC Evol Biol* 16:1–11
- Bentkowski P, Radwan J (2019) Evolution of major histocompatibility complex gene copy number. *PLoS Comput Biol* 15:e1007015
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* 16:363–377
- Biedrzycka A, O'Connor E, Sebastian A, Migalska M, Radwan J, Zając T et al. (2017a) Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evol Biol* 17:159
- Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J (2017b) Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Mol Ecol Res* 17:642–655
- Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176:1035–1047
- Borg ÅA, Pedersen SA, Jensen H, Westerdahl H (2011) Variation in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with different population histories. *Ecol Evol* 1:145–159
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL et al. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39
- Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Mol Biol Evol* 25:1180–1191
- Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L (2010) Adaptive divergence of ancient gene duplicates in the avian MHC class II β . *Mol Biol Evol* 27:2360–2374
- Canal D, Alcaide M, Anmarkrud JA, Potti J (2010) Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Res Notes* 3:236
- Chen LC, Lan H, Sun L, Deng YL, Tang KY, Wan QH (2015) Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Sci Rep* 5:7963
- Cloutier A, Mills JA, Baker AJ (2011) Characterization and locus-specific typing of MHC class I genes in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci. *Immunogenetics* 63:377–394
- Cutrerera AP, Lacey EA (2007) Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (Rodentia: Ctenomyidae). *Immunogenetics* 59:937–948
- Drews A, Strandh M, Råberg L, Westerdahl H (2017) Expression and phylogenetic analyses reveal paralogous lineages of putatively

- classical and non-classical MHC-I genes in three sparrow species (Passer). *BMC Evol Biol* 17:152
- Drews A, Westerdahl H (2019) Not all birds have a single dominantly expressed MHC-I gene: transcription suggests that siskins have many highly expressed MHC-I genes. *Sci Rep* 9:19506
- Edwards SV, Wakeland EK, Potts WK (1995) Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. *Proc Natl Acad Sci USA* 92:12200–12204
- Eimes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y (2015) Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ* 3:e853
- Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M et al. (2006) Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol Lett* 2:543–547
- Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 160:712–716
- Gamer M, Lemon J, Fellows I, Singh P (2012) irr: various coefficients for interrater reliability and agreement. R package ver 84. <https://CRAN.R-project.org/package=irr>
- Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573–582
- Gillingham MA, Courtiol A, Teixeira M, Galan M, Bechet A, Cezilly F (2016) Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species. *J Evol Biol* 29:438–454
- Goebel J, Promerová M, Bonadonna F, McCoy KD, Serbielle C, Strandh M et al. (2017) 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. *BMC Genomics* 18:460
- Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864
- Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131
- He K, Minias P, Dunn PO (2021) Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biol Evol* 13:evaa270
- Hess CM, Edwards SV (2002) The evolution of the Major Histocompatibility Complex in birds: scaling up and taking a genomic approach to the major histocompatibility complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. *Bioscience* 52:423–431
- Högstrand K, Böhme J (1999) Gene conversion can create new MHC alleles. *Immunol Rev* 167:305–317
- Janeway JCA, Travers P, Walport M, Shlomchik MJ (2001) Immunobiology: the immune system in health and disease. Garland Science, New York
- Jarvi SI, Tarr CL, McIntosh CE, Atkinson CT, Fleischer RC (2004) Natural selection of the major histocompatibility complex (Mhc) in Hawaiian honeycreepers (Drepanidinae). *Mol Ecol* 13:2157–2168
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491:444–448
- Kaufman J, Milne S, Gööbel TW (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401:923–925
- Kikkawa EF, Tsuda TT, Sumiyama D, Naruse TK, Fukuda M, Kurita M et al. (2009) Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*). *Immunogenetics* 61:341–352
- Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208–1222
- Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. *PLoS Genet* 4:e1000304
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol Rev* 190:95–122
- Li D, Sun K, Zhao Y, Lin A, Li S, Jiang Y et al. (2017) Polymorphism in the major histocompatibility complex (MHC class II B) genes of the Rufous-backed Bunting (*Emberiza jankowskii*). *PeerJ* 5:e2917
- Li L, Zhou X, Chen X (2011) Characterization and evolution of MHC class II B genes in ardeid birds. *J Mol Evol* 72:474–483
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562–563
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:vev003
- Migalska M, Sebastian A, Radwan J (2019) Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proc Natl Acad Sci* 116:5021–5026
- Miller HC, Lambert DM (2004) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* 56:178–191
- Minias P, Bateson ZW, Whittingham LA, Johnson JA, Oyler-McCance S, Dunn PO (2016) Contrasting evolutionary histories of MHC class I and class II loci in grouse—effects of selection and gene conversion. *Heredity* 116:466–476
- Minias P, Pikus E, Whittingham LA, Dunn PO (2018) A global analysis of selection at the avian MHC. *Evolution* 72:1278–1293
- Minias P, Pikus E, Whittingham LA, Dunn PO (2019) Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biol Evol* 11:17–28
- Morand S, Poulin R (1998) Density, body mass and parasite species richness of terrestrial mammals. *Evol Ecol* 12:717–727
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL et al. (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30:1196–1205
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94:7799–7806
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Ann Rev Genet* 39:121–152
- Nowak MA, Tarczy-Hornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proc Natl Acad Sci* 89:10896–10899
- O'Connor EA, Cornwallis CK, Hasselquist D, Nilsson JÅ, Westerdahl H (2018) The evolution of immunity in relation to colonization and migration. *Nat Ecol Evol* 2:841–849
- O'Connor EA, Strandh M, Hasselquist D, Nilsson JÅ, Westerdahl H (2016) The evolution of highly variable immunity genes across a passerine bird radiation. *Mol Ecol* 25:977–989
- O'Connor EA, Westerdahl H, Burri R, Edwards SV (2019) Avian MHC evolution in the era of genomics: phase 1.0. *Cells* 8:1152
- Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* 11:469–482
- Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218–225

- Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* 401:877–884
- Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 98:13757–13762
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490
- Promerová M, Albrecht T, Bryja J (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics* 61:451–461
- Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J (2020) Advances in the evolutionary understanding of MHC polymorphism. *Trends Genet* 36:298–311
- Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3:217–223
- Revell LJ, Harmon LJ, Collar DC (2008) Phylogenetic signal, evolutionary process, and rate. *Syst Biol* 57:591–601
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542
- Roved J (2019) MHCtools: analysis of MHC data in non-model species. R package ver. 1.3.0. <https://cran.r-project.org/package=MHCtools>
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE et al. (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302
- Sallaberry-Pincheira N, González-Acuña D, Padilla P, Dantas GP, Luna-Jorquera G, Frere E et al. (2016) Contrasting patterns of selection between MHC I and II across populations of Humboldt and Magellanic penguins. *Ecol Evol* 6:7498–7510
- Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 11:1423–1425
- Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* 41:2481–2491
- Saper MA, Bjorkman P, Wiley DC (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol* 219:277–319
- Sebastian A, Herdegen M, Migalska M, Radwan J (2016) AMPLI-SAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol Ecol Res* 16:498–510
- Sepil I, Moghadam HK, Hucard E, Sheldon BC (2012) Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol Biol* 12:68
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:114–114
- Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34:126–129
- Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc R Soc B* 277:979–988
- Spurgin LG, Van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC et al. (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Mol Ecol* 20:5213–5225
- Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PLoS ONE* 9:e100587
- Takahashi K, Rooney AP, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. *J Hered* 91:198–204
- Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc Natl Acad Sci USA* 87:2419–2423
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL (2018) Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 35:773–777
- Winkler DW, Billerman SM, Lovette IJ (2015) Bird families of the world: an invitation to the spectacular diversity of birds. Lynx Edicions, Barcelona
- Witzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Mol Biol Evol* 16:479–490
- Wutzler R, Foerster K, Kempenaers B (2012) MHC class I variation in a natural blue tit population (*Cyanistes caeruleus*). *Genetica* 140:349–364